DOCUMENT RESUME

ED 274 668 TM 860 434

AUTHOR Nenty, H. Johnson
TITLE Cross-Cultural Bias Analysis of Cattell Culture-Fair Intelligence Test.
PUB DATE Apr 86
NOTE 37p.; Paper presented at the Annual Meeting of the American Educational Research Association (67th, San Francisco, CA, April 16-20, 1986).
PUB TYPE Speeches/Conference Papers (150) -- Reports - Research/Technical (143)

EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS Construct Validity; Cross Cultural Studies; Cultural Influences; *Culture Fair Tests; Difficulty Level; Foreign Countries; Indians; Intelligence; *Intelligence Tests; *Item Analysis; Secondary Education; *Test Bias; Test Reliability; *Test Validity; Visual Measures
IDENTIFIERS *Cattell Culture Fair Intelligence Test; Fluid Intelligence; India; Nigeria

ABSTRACT

The Cattell Culture Fair Intelligence Test (CCFIT) was administered to a large sample of American, Nigerian, and Indian adolescents, and item data were examined for cultural bias. The CCFIT was designed to measure fluid intelligence, which is not influenced by cultural differences. Four different item analysis techniques were used to determine cross-cultural validity: Scheuneman's modified chi-square procedure (SSX2); Rudner and Convey's TID-45 degrees item difficulty p-value; one-parameter latent trait Rasch model; and Cochran's Test Method (CTX2) using chi-square. Results indicated a high inter-method correlation, strengthening the convergent validity of the methods used. CCFIT's claim of measuring a single second-order intelligence factor, fluid intelligence, was supported. However, 59 percent of the items were identified as biased, strongly suggesting the CCFIT is not free of cultural bias. Hence, it may not have universal validity as a measure of fluid intelligence. (GDC)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED274668

CROSS-CULTURAL BIAS ANALYSIS OF CATTELL CULTURE-FAIR INTELLIGENCE TEST

BY:

H. JOHNSON NENTY
FACULTY OF EDUCATION
UNIVERSITY OF CALABAR
NIGERIA

PAPER PRESENTED AT THE ANNUAL MEETING
OF THE AMERICAN EDUCATIONAL RESEARCH ASSOCIATION
SAN FRANSISCO, CALIFORNIA, APRIL 17, 1986

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

✓ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

H. Johnson Nenty

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

TM 860434

BEST COPY AVAILABLE

CROSS-CULTURE BIAS ANALYSIS OF CATTELL CULTURE-FAIR INTELLIGENCE TEST

by:

H. Johnson Nenty
University of Calabar
Nigeria

ABSTRACT

The Cattell Culture Fair Intelligence Test (CCFIT) item data from three mutually remote and culturally disparate groups (600 Americans, 231 Indians, and 803 Nigerians) were analysed for bias using four techniques in order to determine the cross-cultural validity of the test. The CCFIT claims to measure a single second-order intelligence factor--fluid intelligence —, and to be indifferent to culture-related factors that might influence examinee's responses to its items. Following from these claims, every examinee's total score should indicate his or her standing on a single, and universally common, continuum that defines fluid intelligence. The high inter-method correlation found and their correspondingly high ratios of agreement strengthened the convergent validity of the methods used. Evidence were found to support results of previous studies in arguing for the first CCFIT claim, but the large proportion (5%) of its items identified as biased strongly suggest that CCFIT might not be free from significant influence of culture related factors, and hence may not have universal validity as a measure of fluid intelligence.

INTRODUCTION

Recent years have witnessed a growing interest in cross-cultural differences (Drenth & Van der Flier, 1976), and this has given rise to a "tremendous expansion in cross-cultural research" (Manaster & Havighurst, 1972, p. ix). According to Manaster and Havighurst (1972), such research has resulted in the collection of very valuable information from the comparison of the behaviour, beliefs, and attitude of subgroups within complex societies. These researchers believe that "the extension of these studies to include groups

in more than one country promises to provide even greater insight into the nature of man as a species" (p. ix). The extent to which such information is "valuable" depends on the answers to some very important theoretical questions such as "Do our explanatory constructs have enough generality?", "Are the Western psychological concepts able to describe and explain non-Western behaviour in a satisfactory way?", "Are the tests (used in such studies) themselves biased or are the interpretation and use what is biased?" (Massad, 1978, p. 3). Most cross-cultural studies are done with little or no concern about the issues raised by these questions, and such studies often result in findings that lead to erroneous conclusions and generalizations. For the results of the "extension or these studies to include groups in more than one country" to be of any scientific value, answers to these questions must first be sought for and found.

Kidd (1962) pointed out that "valid comparison of scientific findings among different cultures and ethnic groups" is only possible with a test consisting of "truly culture-fair items" (p. 352). Though most experts in cross-cultural psychology generally deny the possibility of a culture-free test, it is believed that with modern measurement techniques, test could be designed, and administered, to minimize and keep constant the influence of cultural experience on examinees' responses. These tests are designed to "avoid relationships depending on anything more than immediately given fundaments that are equally

unfamiliar to all" (Catteli & Horn, 1978, p. 141). Through construction, "extrinsic" cognitive factors that might differentially influence examinee's responses, like extent of examinees' familiarity with the language of the test, and level of experience with the test stimuli (content), are controlled for, and through administration, extrinsic non-cognitive factors like motivation to perform well, test sophistication, speed, rapport with the examiners, understanding of test instructions and the strategy of answering the test questions, are controlled or kept constant for all examinees.

Any of these extrinsic factors may differ significantly with culture, and because different cultures foster the development of different values, skills, and knowledge, responses of examinees from different cultures to the same test item may reflect these differences, rather than their differences on the underlying ability in the attribute which the test was designed to measure—" intrinsic" factor. In other words, in the presence of extrinsic factors, the same test item may elicit different responses from examinees from different cultures even though they may have the same ability on the trait which the test was designed to measure. Under such condition, an examinee's score cannot be interpreted exclusively as his or her level of attainment on the characteristics one was trying to measure. Except with a proven culture-fair test for the cultures

involved, this makes it practically impossible to measure objectively a person's true standing on some attribute of interest and thus illogical to compare scores on the same "test" across cultures or subcultures. On this basis Green and Draper (1972) noted that "a systematic examination for cultural bias of the major published ability ... tests now in use ... is long overdue" (p. 6). The significant presence of extrinsic factors like school training and life or cultural experiences in traditional intelligence test like Standford-Binet (Cattell, 1940) disqualifies them from any cross-cultural application.

Massad (1978) listed three requirements that a test should meet before it could be regarded as a valid instrument for cross-cultural comparison. These requirements are "(1) functional equivalence (the test has to measure the same attribute on a similar scale); (2) score equivalence (the test has to measure this attribute on a similar scale); (3) item equivalence (the same requirement as score equivalence, but applied on the item level)" (p. 3). Following from these requirements, the validation of a test for cross-cultural research should proceed by checking (1) how well the test measures, on the same scale, the same attribute for the different cultural groups concerned; and (2) how similar is the behaviour of each of the items in measuring on the same scale, the attribute for each of the groups.

BACKGROUND

Item-bias Detection Methods.

Many methods are available for use in testing for one or both of the above requirements for any test intended for cross-cultural application. Such methods have been reviewed by Jensen (1980) and Merz and Rudner (1978) and compared by Ironson & Subkoviak (1979) and Rudner et. al. (1980). Scheuneman's (1979) modified chi-square procedure, one of these methods (SSX^2), defines an item as unbiased if, for all persons of equal ability (i.e., equal total score on a test containing the item) the probability of a correct response is the same regardless of each person's cultural or ethnic group membership. With this procedure, each item is separately tested for bias by first establishing a number of ability or total score intervals based on the total score on the rest.

Another of the most frequently used methods for examining items for bias is the comparison, across each pair of groups, of each item difficulty (p-value) as defined by the classical test theory. An item is biased if its p-value, or a transformation of this value, differs significantly across the groups under comparison. One version of this approach, the TID - 45^0 (Rudner & Convey, 1978; Rudner et. al., 1980) entails standardizing the p-values to within-group z-value.

The one-parameter latent trait model (ICC-1) attributed to Rasch (1960), stipulates a probabilistic relationship

between observable item and person scores and the latent item and ability parameter assumed to underline the result of a person-by-item encounter during testing. The model holds that, given a undimensional test, the probability of an examinee succeeding on any item is wholly determined by his/her ability on the trait that is being measured, and by the difficulty of the item. It assumes the absence of guessing and constant discrimination for all the items of the test. The model implies that "the estimate of the item difficulty parameter will not vary significantly over different samples ... [but the] item parameter will not be sample-invariant where there is culture bias which differentially affects the item probability". Following from this "if a test includes items which differ in cultural loadings, the special conditions required for item parameter invariance may be difficult to obtain" (Whitely & Dawis, 1974, p. 175-176). To test for bias with this method, one procedure is to calibrate the items separately for each cultural group, and then compare, for each item, the Rasch difficulty parameter estimates across groups (Henty and Dinero, 1981).

The Cochran's Test Method (CTX^2) is one of the six methods suggested by Marascuilo and Slaughter (1981) for identifying item bias based on chi-square statistics. It is recommended for use in the case in which the belief is that the differences between the probabilities of success

for subjects in each group are equal for all ability intervals. Hence the null hypothesis associated with this procedure is that there is no difference in the probabilities of success across all the ability group. A rejection of this hypothesis is an indication of differential difficulty.

Studies done to establish the validity of these methods have demonstrated that they do not select items at random (Ironson & Subkoviak, 1979; Rudner and Convey, 1978). In an artificially created situation, the amount of bias detected by SSX^2, TID, and ICC-1 methods correlated .73, .68 and .60 respectively with the actual amount of bias generated (Rudner et. al., 1980). Rudner and Convey (1978) found three methods: item characteristic curve with three parameters (ICC-3); SSX^2, and TID methods out of seven to be most promising, and observed the intercorrelation of their detected amount of bias to be .67 between SSX^2 and ICC-3, .59 between SSX^2 and TID, and .31 between ICC-3 and TID. SSX^2 have recently been critizied (Baker, 1981; Marasculo and Slaughter, 1981) and related procedures including CTX^2 has been suggested (Marascuilo & Slaughter, 1981).

Cattell Culture Fair Intelligence Test (CCFIT)

According to Cattell's theory of intelligence, all intellectual activities are sustained by two general factors, one fluid (g_f) and the other crystallized (g_c). While fluid intelligence is assumed due to biological factors, crystallized intelligence supposedly develops through experience with culture. In other words, while the latter

is culture dependent, "innate mental ability"--fluid intelligence— is not. Thus differences in fluid intelligence might exist between individuals but not necessarily between cultures as might be the case with crystallized intelligence. From the moment g_f was recognized as a distinct factor, Cattell and his colleagues began "to search for a more 'saturated' measure to define it [operationally], and ... [the result] turned out, almost as a by-product, to have promising properties as culture-fair tests" (Cattell, 1971, p. 82). Since then, "culture-fair intelligence tests... have become the practical test of expression of fluid general ability factor" (Cattell, 1971 p. 78). The Cattell Culture Fair Intelligence Test (CCFIT) is assumed to measure, for every culture, a single second-order intelligence factor-fluid intelligence — (Cattell, 1971, p. 78), and to be indifferent to culture-related factors that might influence examinees' responses to its items (Cattell, 1968, p. 61). Following from both assumptions, a person's total score on CCFIT should be an indication of his or her standing on a single continuum that defines fluid intelligence and each item of the test should elicit the same response from all persons with the same level of fluid intelligence irrespective of cultural experiences. Examinee's variability in such responses should be due only to the individual differences in their fluid intelligence and not to other factors on which they might differ, and the test variance can be explained by

the same theoretical construct for each cultural group.

Many studies have been done to test the two major claims of CCFIT: a high saturation on general ability factor and a relative independence from cultural experiences. Generally, the first claim has been supported by most related studies, but many investigators using the test reported limited success on the second (Anastasi, 1968; Kidd, 1962). It was generally found that with the CCFIT, norms tended to remain unchanged in cultures moderately similar to that which the test were developed; in other cultures, however, performance fell considerably below the original norm (Anastasi, 1968, p. 246-47). But the methods often used in these studies, for example, comparison of group means, analysis of variance, and factor analysis are inadequate as methods of detecting bias (Lord & Novick, 1968).

METHOD

Sample

The sample used in this study included 600 American, 231 Indian, and 803 Nigerian students[1]. The subjects in each of the three cultural groups were mainly 13-15 years old boys and girls, and were comparable in the number of years of normal schooling. Those in the Indian and Nigerian samples were able to communicate effectively in English. Subject were members of existing grades or classes in schools that

[1]The help rendered by Dr. G.B. Shah and Patrick J. Udoudo in the collection of test data in India and Nigeria is gratefully acknowledged.

participated in the study.

Instrument

Scale 2, Form A of the CCFIT was used. The test is reported to have a Spearman-Brown reliability of .79, a Cronbach's alpha of .77 and a KR21 of .81 (IPAT, 1973). Knapp (1960) reported a split-half (even-odd) reliability coefficient of .90 and .92 under speeded conditions for Mexican and American samples respectively, .71 and .89 were reported under power conditions. IPAT (1973) reported a construct validity of .85 for the test, and also indicated that in a series of factor analytic studies correlations between Form A and B and the general ability factor, g, were found to range from .73 to .83.

The CCFIT is a paper and pencil test which makes use of abstract geometric forms. It consists of four subtests all measuring a single factor-fluid intelligence. The first subtests (12 items) requires the examinee to select, from among five choices, the one geometric form that best continues a progressive series. In the second subtest (14 items), the examinee is required to select from among five geometric forms one which is different from the other four. The third subtest (12 items) requires the examinee to complete a matrix presented at the left of each row by choosing one among five geometric forms presented at the right. The last subtest (8 items) requires the examinee to select from five choices provided in the right of each row the one which duplicates the conditions of the geometric figures given on the left of the row.

Through its method of construction, the CCFIT purports to rule out, or hold constant, the influence of contaminating cognitive factors such as language and familiarity with test content. In this study, attempts were made to control additionally for possible non-cognitive contaminating factors like interest and motivation, speed, and test sophistication.

Analysis of Data

The mean, standard deviation and the Kuder-Richardson 20 reliability coefficient for each of the three cultural groups were calculated. To ascertain the extent to which the data meet the undimensional assumption underlying all the bias-detection procedures, a factor analysis of the data for each of the cultural groups was done. Given the disadvantage that both phi and tetrachoric coefficients have limitations which influence the results of such analysis when dichotomous scores are factored (Lord & Novick, 1968, p. 382), a principal factor analysis was done on subjects' subtests score. The SPSS program PA2 was used for the analysis (Snepard et. al., 1981).

Each of the four methods were applied to calculate the item-bias indices for each pair of the three cultural groups - American/Indian, American/Nigeria, and Indian/Nigeria. For the Scheuneman and Cochran methods five total-score, intervals were selected such that the mean scores for the three groups within each total-score interval were statistically equivalent ($p. < .01$). The cell frequencies for six of

the items were less than ten for two of the three comparisons, therefore four intervals were used for one, and three intervals for five of these items. While the bias index for Cochran test was the Z_c - value, that of Scheuneman method was the mean-square value got by dividing the signed chi-square value by the number of degrees of freedom.

With the TID-45^0 method, the standard deviation of distances of item-points to a 45^0 line was a computed, and biased items were classified as those deviating by greater than 1.5 standard deviation units from the line (Mudner, 1977). Finally, for the ICC-1 method, item difficulty parameter estimates and their corresponding standard error for each group was gotten using BICAL (Wright & Mead, 1978) computer program. A t-test of fit to the model was done for each item, and for each of the three groups. With these statistics, a pairwise comparison of the difficulty parameter was done and the bias index was the resulting t-value. An alpha level of .05 was used to select items as biased with ICC-1, and Cochran methods, but .30 was used with Scheuneman method (Scheuneman, 1979).

RESULTS

The raw score means and standard deviations for each cultural group are presented on Table 1. The Kuder-Richardson 20 reliability coefficients were .82, .72, and .75 for the American, Indian, and Nigerian groups respectively.

In the factor analysis, the first unrotated factor accounted for 90%, 78% and 89% of the total variance for the three respective groups. No non-fitting item ($p<.05$) to the Rasch model was observed for any of the three groups. Based on these, the test could be assumed to satisfy the unidimensional requirement.

Insert Table 1 about here

For the American/Indian comparison, at least three of the four methods flagged the same twenty-one items as biased, eleven for the Americans, and ten for the Indians (see Tables 2 through 5).

Insert Tables 2 through 5 about here

All the six items identified by these methods as biased in Subtest 1 were in favor of the American group, similarly, the five items identified as biased in Subtest 4 were all in favor of the Indian group.

In the American/Nigerian comparison, at least three of the four methods identified the same seventeen items as being biased, nine for the Americans and eight for the Nigerian. While in the Indian/Nigerian comparison, at least three of the four techniques flagged the same 15 items as biased seven for the Indians, and eight for the Nigerians (see Table 2 through 5).

A close examination of the converging items revealed some generalizable patterns between the four subtests and the

three cultural groups. Of the 15 items which tended to favor the Americans, seven were from Subtest 1, and only one was from Subtest 4, whereas for the 12 items which tended not to favor them, five were from Subtest 4 and only one from the first subtest. On the other hand, for the Indians, five of the ten items which tended to favor them were from the fourth subtest, and one was from subtest 1, while seven of the 13 items biased against them were from Subtest 1 and none from Subtest 4. Infact, no item in Subtest 4 favored any other group when compared with the Indian; and no item in Subtest 1 favored the Indian group when compared with any another group.

For the Nigerian group the relationship tended to be less obvious, but compared with this group only, one item in Subtest 3 was biased in favor of the Indian and the American group respectively. Out of the 14 items that tended to be biased against the Nigerians, only one was from subtest 3. No other group had as many items which tended to favor them, and as few items which did not favor them in Subtest 3 as the Nigerians.

Using a 2 x 3 contingency table for each subtest and the number of items "for" and "against" each of the three groups, a chi-square analysis gave a significant ($x^2 = 11.52$, df - 2, $p < .01$) evidence of overall bias of Subtest 1 for the American group, and a similar evidence ($X^2 = 7.63$, df = 2, $p < .05$) of bias of Subtest 4 for the Indian group.

Though there was a lack of statistical evidence (X^2 = 2.52), substantively, Subtest 3 tended to favor the Nigerian group more than any of the other groups. The same tendency, though still less pronounced (X^2 = 1.91), was observed of subtest 2, for the American group. Only seven of the 46 items in the test were not identified by any method as being significantly biased in any of the three group comparisons (see Table 2 through 5).

Generally, in a relative sense, Subtests 1 and 2 tended to favor the American group, Subtest 4, the Indian groups, and Subtest 3, the Nigerian group. On the other hand, Subtest 4 tended not to favor the American group, Subtest 1, the Indian group, and Subtests 2 and 4, the Nigerian group. Subtest 3 had the least proportion of biased items, and did not show any sign of significant bias against any of the three cultural groups. Similar trend was observed for the American and Nigerian groups when Nenty and Dinero (1981), using the Rasch model only, analysed the fairness of CCFIT for these two groups.

In view of the observed trend of relationship between the four CCFIT subtests and the three cultural groups, an attempt was made to further analyse the basic general skill requirement of these subtests. Subtests 1 and 2 both consisted of basically classification items (which of these five figures, given the first three of the left, belongs?, or does not belong?). The stimuli is not different from that of any series or classification item in an English

grammar, or a mathematics test, except here the language is figural. Each of these is solved by first establishing a pattern, or trend, and then identifying, through logical comparisons, one out of five alternatives that fits, or fails to fit the established pattern.

The third subtest consisted of progressive matrices with varying matrix patterns. Progressive matrizes are said to be among the best tests of eduction (Jensen, 1980). It is more spatial than the first two subtests but like them, solution to its items involved an initial establishment of a pattern of relationship between the different entries of the two-or-three dimensional sequence to the left of each row. Then through extrapolating and analogic reasoning one out of five alternatives that most logically completes the matrix on the left is chosen. Subtest 4 consisted of topology items, and these are the most spatial of all the items in the test. An item is solved by first establishing the position of a dot with relation to the geometric figures in the square to the left of each row. In the five alternatives given on the right of each row, the geometric figures used in each item have been moved about, without any trend or pattern, in a two-dimensional space. Basically by logical comparisons and eliminations, one of these five alternatives for which the relative position of the dot could be maintained is chosen.

The mean of the bias indices generated by each method for each group comparison (Table 6) showed large

Insert Table 6 about here

difference between the biased, unbiased, and all the items. On the average, the means of the indices for the biased items were about twice as large as that for all the items, and more than four times larger than for the unbiased items. This magnitude of difference implies that on a continuum reflecting degree of bias, only extremely biased items were identified. A Pearson correlation and a ratio of agreement (Shepard et. al., 1981) analyses of the bias indices generated showed, for each of the three comparisons, strikingly high coefficients of correlation and correspondingly high ratios of agreement between the four item-bias detection methods (Table 7 and 8). All the methods agreed on the direction of bias of any item identified as being significantly biased.

Insert Table 7, and 8 about here

DISCUSSION

Though the test K-R 20 reliability index found for each cultural group in this study was not significantly different from most other published data for CCFIT administered under power conditions, there was a significant difference ($p < .001$) between these internal consistence estimates for the Indian and Nigerian groups, and that for the American group. Such differences in the test reliability between the groups indicated that the test variance attributable to random error was substantially larger for the Indian and the

Nigerian groups than for the American group. Thus, the test yielded more inaccurate scores for the Indian and the Nigerian groups than it did for the American group (Green & Draper, 1972). This condition typifies Green and Draper's (1972) "Type A" bias—a bias signified by inequality in test reliability, a condition where bias exists in an instrument which though measuring the one main factor for all groups involved, shows significantly unequal item loadings on this factor for each of these groups. Such condition is also an example of an instrument which might meet the first of Massad's (1978) three requirements for a valid instrument for cross-cultural comparison, but might fail to meet the remaining two.

The inter-method correlations found in this study are comparable with those reported by Rudner and his colleagues (1980) who used Monte Carlo generated item response data. They are, however, significantly larger than related indices reported by Ironson (1979) and Rudner and Convey (1978) for real data. The high degree of agreement observed in this study might stem from the fact that the methods used were all found to be sensitive to bias caused by extreme differences in item difficulty (Rudner et. al., 1980).

The CCFIT, by construction, assumes to measure, for every culture, a single intelligence factor, and results of related studies with different cultures tend to confirm this. In this study, both the large percentage of test variance accounted for by the first factor, and the absence of a non-fitting item to the Rasch model for each of the three groups

support the findings of these related studies in arguing for the validity of this assumption. With this condition, group differences in ability distributions may not create artifactual instances of bias when item parameter is sample invariant (Lord and Novick, 1968) like the parameter estimated with ICC-1 method (Wright & Mead, 1978). The large number of items identified by techniques which differ in their susceptibility to artifactual causes of bias coupled with the high indices of agreement between these methods strongly suggest that:

(1) detection of bias was not a random process, but there was a consistent factor of bias upon which these techniques converge; and (2) it is less likely that such a systematic factor is artifactual. Since all the methods applied are related to item difficulty, significant differential difficulty might be the source of bias upon which these methods converge.

A high degree of agreements between test item bias detection methods is an indication of the amount of bias in the test (Shepard et. al., 1981). Twenty seven out of the 46 items of CCFIT were identified by at least three methods as being most likely biased. As the number of items found to be biased in a test might give an indication of the overall amount of bias in that test, it is likely that for every culture, CCFIT is not equally free from bias. This large number of items that tended to be bias provides a very compelling evidence against the assumption

that the test is equally free from cultural factors the influence of which might significantly affect the outcome when it is used for cross-cultural measurement and comparison.

The Rasch model, used as the only bias-detection method, identified 22 CCFIT items as unbiased for both the American and Nigerian group involved (Nenty & Dinero, 1981). In that study, bias detection analysis was not done for 13 CCFIT items which showed significant between-groups fit statistics. In this study using four methods, a total of 29 items are identified as unbiased for the same two cultural groups. Seventeen of these are shared with the 22 Rasch-detected unbiased items identified in the former study. Thus an increase in the number of bias-detection methods has reduced the number of items identified as being unbiased. Furthermore, as a result of the three pairwise comparisons, these four methods have identified 19 items as being unbiased for all the three groups involved. In this case, an increase in the number of cultural groups have significently reduced the number of items identified as unbiased. If this trend continues, sufficient increase in the number of cultural groups studied might result in the identification of all items as biased. Therefore there might not be a set of items which could be said to be fair to all cultural groups. It might then be safer to define culture fairness in a relative, rather than in a universal sense.

The analyses of the skills involved in each of the subtests showed that though solution to their items

involves initially the eduction of relationships, and then some logical comparisons, there are some differences in the third level skill requirements by some of the subtests. The trend of relationship observed between these subtests and the three cultural groups might stem from the differences that might exist between these groups on the basic intrinsic skills involved in each of the subtests. Mostly based on the presence of such unique skills and the impossibility of standardizing all extrinsic factors, most cross-cultural psychologists deny the existence of culture-fair test (Cronback & Drenth, 1972; Massad, 1978). They maintain that cognitive functioning is an adaptation to specific cultural and ecological demands, and should be defined in terms of such requirements. They argue against generalizing from test scores to non-test behaviour for all groups without first explicitly validating the test for each of the groups involved. Thus they insist on less transfer of test validity across cultural lines, and less generality in the interpretation of test performance.

Although this study found bias to be a significant factor for the three cultural gruups used, and though the presence of bias in test items distorts all classical item parameters, as well as group means, the study did not provide evidence of the impact of such bias on the test results. It did not estimates how much, if any, of the differences in mean scores that was observed (see Table 1)

between the cultural groups could be accounted for by the 27 items identified as being significantly biased in the test. The study was concerned not with the issues of determining which cultural gruop is intellectually superior or inferior through comparison of group means, or the contributions of heredity and environment to intellectual development, but with the very basic question of the fairness, or cross-cultural validity, of one of the instruments most often used to obtain the scores on which decisions on such issues are based.

CONCLUSIONS

Given the three cultural groups, and the four item-bias methods used in this study, 27 out of the 46 items of CCFIT tended to be relatively biased. There was evidence that the test measured a single main factor for each of the groups, but that such a large proportion (59%) of its items was identified as biased and with a high degree of agreement between the different methods strongly suggest that CCFIT might not be equally free from culture-related factors for all cultures. This study was not concerned with the universal validity of fluid intelligence, but based on its results CCFIT designed as a measure of this construct is not free from significant cultural influences. As the number of items that could be said to be fair diminished with increase in the number of cultural groups studied, there might not be a set of items that provides a universal definition of intelligence.

REFERENCES

Anastasi, A. Psychological testing (1968) New York: Macmillan Publishing Col, Inc.

Baker, Frank B. (1981). Criticism of Scheuneman's item bias Technique. Journal of Educational Measurement, 18, 59-62.

Cattell, R.B. (1968). Are 10 tests intelligence?" Psychology Today, March, 56-62.

Cattell, R.B. (1971). Abilities: Their structures growth and Action. Boston: Houghton Mifflin.

Cattell, R.B., & Horn, J.L.A. (1978). A check on the theory of fluid and crystalized intelligence with description of new subtest designs. Journal of Educational Measurement, 15, 139-164.

Cronbach, L.J., & Drenth, P.J.D. (1972). Summary and Commentary. In L.J. Cronbach & P.J.D. Drenth (Eds.) Mental Test and Cultural Adaptation. The Hague: Houton Publishers.

Drenth, P.J.D. (1977). The use of intelligence tests in developing countries. In Ype H. Poortinga (Ed.) Basic Problems in Cross-Cultural Psychology. America Amsterdam: Swets & Zeithlinger, B.V.

Drenth, P.J.D. & Van der Flier, H. (1976). Cultural differences and Comparability of test scores. International Review of Applied Psychology, 25, 137-144.

Green, D.R., & Draper, J.F. (1972). Exploratory studies of bias in achievement tests. Monterey, California: CTS/ McGraw. (ERIC Document Reproduction Service No. ED 070 794).

Institute of Personality and Ability Testing (IPAT) (1973). Measuring intelligence with the culture fair tests. Manual for scales 2 and 3. Chicago: IPAT.

Ironson, Gail H., & Subkoviak, Michael J.A. (1979). A comparison of several methods of assessing item bias. Journal of Educational Measurement, 16, 209-225.

Jensen, Arthur R. (1980). Bias in Mental testing. New York: The Free Press.

Kidd, A.H. (1962). The culture fair aspects of Cattell's test of g: Culture-free The Journal of Genetic Psychology, 101, 343-362.

Knapp, R.S. (1960)."The effects of time limits on the intelligence test performance of Mexican and American subjects." Journal of Educational Psychology, 51, 14-20.

Manaster, G.J., & Havighurst, R.J. (1972). Cross-National Research: Social-Psychological Methods and Problems. Boston: Houghton Mifflin Co.

Marascuilo, L.A., & Slaughter, P.E. (1981). Statistical Procedures for identifying possible scources of item bias based on X^2 statistics. Journal of Educationmal Measurement, 18, 229-249.

Massad, C.E. (1978). International Scisne. Measurement News, National Council of Measurement in Education, 2 3.

Merz, W.R. & Rudner, L.M. (1978). Bias in testing: A Presentation of selected methods. Paper presented at the annual meeting of the American Educational Research Association, Toronto.

Nenty, H.J. & Dinero, T.E. (1981). A Cross-Cultural Analysis of the Fairness of Cattell Culture Fair Intelligence Test Using the Rasch Model. Applied Psychological Measurement 5, 355-368.

Rasch, G. (1960). Probabilistic Models for some intelligence and Attainment Tests. Copenhagen: Danish Institute for Educational Research.

Rudner, L.M. (1977). An approach to biased item identification using latent trait measurement theory. Paper presented at the annual meeting of the American Educational Research Association, Toronto, Canada.

Rudner, Lawrence M., Getson, Pamela R., & Night, David L. (1980). A Monte Carlo Comparison of seven biased item detection techniques. Journal of Educational Measurement 17, 1-10.

Scheuneman, Janice, (1979). A method of assessing bias in test items. Journal of Educational Measurement, 16, 143-153.

Shepard, Lorries, Camilli, Gregory, & Averill, Marylyn, (1981) Comparison of procedures for tecting test - item bias with both internal and external ability criteria. Journal of Educational Statistics, 6. 317-375.

Whitely, S.E., & Davis, R.V. (1974). The nature of objectivity with the Rasch Model. Journal of Educational Measurement, 11, 163-178.

Wright, B.D., Mead, R.J. (1978). BICAL: Calibrating rating scales with the Rasch Model. Research Memorandum No 23A. Chicago: University of Chicago, Department of Education, Statistical Laboratory.

TABLE 1

Mean and Standard Deviation
For Each Cultural Group

Cultural Group	n	x	SD
American	600	31.18	6.18
Indian	231	29.95	5.43
Nigerian	803	30.15	5.68

TABLE 2

Items in Subtest 1 Identified as Biased by
Each Method and for Each Comparison Group

Group Comparison	Bias Method	Item											
		1	2	3	4	5	6	7	8	9	10	11	12
American/ Indian	SSX^2	-	-	-	-	-	A[a]	-	-	A	A	-	A
	$TID\text{-}45^0$	-	-	-	-	-	A	A	A	A	A	A	A
	ICC-1	-	-	A	-	-	A	A	A	A	-	-	A
	CTX^2	-	-	A	-	A	A	A	A	A	A	A	A
American/ Nigerian	$SS\ X^2$	-	-	-	-	-	-	A	A	-	N	A	A
	$TID\text{-}45^0$	-	-	-	-	-	-	A	A	A	-	A	A
	ICC-1	-	-	N	-	-	A	-	A	A	N	-	-
	CTX^2	-	-	N	-	-	A	A	A	A	N	A	A
	$SS\ X^2$	-	-	-	-	-	N	-	-	N	N	-	-
	$TID\ 45^0$	-	-	N	-	-	N	-	-	N	N	-	-
	ICC-1	N	-	N	-	N	N	-	-	N	N	-	-
	$CT\ X^2$	-	-	N	-	N	N	N	-	N	N	-	-

A = Item tends to be biased for the American group
I = Item tends to be biased for the Indian group
N = Item tends to be biased for the Nigerian group

TABLE 3

Items in Subtest 2 Identified as Biased by
Each Method and For Each Comparison Group

Group Comparison	Bias Method	Item 1	2	3	4	5	6	7	8	9	10	11	12	13	14
American/ Indian	SSX^2	-	-	-	A[a]	A	-	I	A	-	I	A	-	A	-
	$TID\text{-}45^0$	-	I	-	A	A	-	I	-	-	I	A	-	-	I
	ICC-1	-	I	A	-	-	-	I	-	-	I	-	I	-	I
	CTX^2	-	I	A	A	A	-	I	A	-	I	A	-	-	-
American/ Nigerian	SSX^2	-	-	-	-	-	-	-	A	-	N	A	-	-	N
	$TID\text{-}45^0$	-	-	-	-	-	-	-	A	-	N	A	-	-	-
	ICC-1	-	A	A	-	-	-	-	A	A	N	A	N	-	N
	CTX^2	-	A	A	-	-	A	-	A	A	N	A	N	-	-
Indian/ Nigerian	SSX^2	-	-	-	N	-	-	I	-	-	I	-	-	-	-
	$TID\text{-}45^0$	-	I	-	N	-	-	I	-	-	I	-	-	-	-
	ICC-1	-	I	-	N	-	-	I	-	-	I	-	-	-	-
	CTX^2	-	I	-	N	N	-	I	-	-	I	-	-	-	-

[a] A = Item tends to be biased for the American group
I = Item tends to be biased for the Indian group
N = Item tends to be biased for the Nigerian group

TABLE 4

Items in Subtest 3 Identified as Biased by Each method and for Each Comparison Group

Group Comparison	Bias Method	Item 1	2	3	4	5	6	7	8	9	10	11	12
American/ Indian	SSX^2	-	-	-	A[a]	-	-	-	-	I	-	A	-
	$TID\text{-}45^0$	-	-	-	A	-	-	-	-	I	-	A	I
	ICC-1	A	-	A	A	-	-	-	I	I	-	-	I
	CTX^2	A	-	-	A	-	-	-	I	-	-	A	I
American/ Nigerian	SSX^2	-	-	-	-	-	-	-	N	N	A	-	-
	$TID\text{-}45^0$	-	-	-	N	-	-	-	N	N	A	-	-
	ICC-1	-	-	-	N	-	N	N	N	-	A	-	-
	CTX^2	-	-	-	N	-	N	-	N	N	A	-	-
Indian/ Nigerian	SSX^2	-	-	-	N	-	-	-	-	-	-	N	I
	$TID\text{-}45^0$	-	-	-	N	-	N	-	-	-	-	N	I
	ICC-1	N	-	-	N	-	N	-		-	I	N	I
	CTX^2	-	-	-	N	-	N	-	-	-	I	N	I

[a] A = Item tends to be biased for the American group
I = Item tends to be biased for the Indian group
N = Item tends to be biased for the Nigerian group

TABLE 5

Items in Subtest 4 Identified on Biased by Each Method and for Each Comparison Group

Group Comparison	Bias Method	Item 1	2	3	4	5	6	7	8
American/ Indian	SSX^2	I[a]	I	I	I	I	-	A	-
	$TID-45^0$	I	I	I	I	I	-	-	-
	$ICC-1$	I	I	I	I	I	-	-	I
	CTX^2	I	I	I	I	I	-	-	-
American/ Nigerian	SSX^2	N	N	A	-	N	-	-	-
	$TID-45^0$	N	N	A	-	N	-	-	-
	$ICC-1$	N	N	A	-	-	-	-	-
	CTX^2	N	N	A	-	N	-	N	-
Indian/ Nigerian	SSX^2	-	I	I	I	-	-	N	-
	$TID-45^0$	I	I	I	I	-	-	-	-
	$ICC-1$	-	I	I	I	I	-	-	-
	CTX^2	-	I	I	I	I	-	-	-

[a] A = Item tends to be biased for the American group
I = Item tends to be biased for the Indian group
N = Item tends to be biased for the Nigeria group.

TABLE 6

Average Item-Bias Indices for Biased and Unbiased Items by Method and Group Comparison

Bias Method[b]	Items	American/Indian N	American/Indian Mean	American/Nigerian N	American/Nigerian Mean	Indian/Nigerian N	Indian/Nigerian Mean
1.	Biased	19	2.82(1.60)	15	2.78(1.57)	13	3.46(1.02)
	Unbiased	27	.41(.35)	31	.40(.33)	33	.63(.41)
	Total	46	1.41(.62)	46	1.21(1.99)	46	1.43(2.39)
2.	Biased	23	2.36(1.01)	15	1.65(.50)	17	2.32(.90)
	Unbiased	23	.67(.35)	31	.47(.29)	29	.60(.32)
	Total	46	1.52(1.13)	46	.65(.67)	46	1.24(1.03)
3.	Biased	25	4.29(2.21)	20	3.64(1.42)	20	4.30(1.75)
	Unbiased	21	.93(.59)	16	.93(.52)	26	.95(.81)
	Total	46	2.61(2.33)	46	2.46(1.75)	46	2.26(2.07)
4.	Biased	28	4.03(1.64)	26	3.50(1.33)	20	4.18(1.75)
	Unbiased	18	.79(.61)	20	.99(.54)	26	.80(.50)
	Total	46	2.40(2.14)	46	2.35(1.64)	46	2.12(2.03)

[a]The bias indices indicate magnitude and not direction of bias.

[b]1 = Signed Scheuman Chi-square; 2 = TID-45[0]: 3 = ICC-1: 4= Cochren Test (chi-square).

[c]Number in parantheses represents standard deviation.

TABLE 7

Ratios of Agreement of Bias Indices
by Group Comparison and Method

Group Comparison	Bias Method	No. of Biased Items	SSX^2 1	TID-45^0 2	ICC-1 3	CT X^2 4
American/ Indian	1	19	-	15/19	12/19	17/19
	2	23	16/23	-	16/23	20/23
	3	25	12/25	17/25	-	20/25
	4	28	17/28	20/28	20/28	-
American/ Nigerian	1	15	-	12/15	14/15	14/15
	2	15	12/15	-	14/15	15/15
	3	28	14/28	14/28	-	24/28
	4	26	14/26	15/26	24/28	-
Indian/ Nigerian	1	13	-	12/13	12/13	13/13
	2	17	12/17	-	15/17	15/17
	3	20	12/20	15/20	-	18/20
	4	20	12/20	16/20	18/20	-

Note: The table is better read by rows, i.e., given the number of biased items identified by the method, how many of these were also identified as biased by the column method.

TABLE 6

Average Agreement Ratios, and Coefficients of Correlation of Bias Indices by Method and Group Comparison

Group Comparison	Bias Method	SSX^2 1	TID-45° 2	ICC-1 3	CTX^2- 4
American/ Indian	1	-	.86[a]	.85	.90
	2	.76	-	.87	.92
	3	.56	.73	-	.94
	4	.75	.79	.76	-
American/ Nigerian	1	-	.86	.77	.81
	2	.80	-	.94	.96
	3	.72	.72	-	.98
	4	.74	.79	.86	-
Indian/ Nigerian	1	-	.84	.83	.87
	2	.82	-	.92	.96
	3	.71	.82	-	.95
	4	.83	.87	.90	-
Overall Average	1	-	.85	.81	.86
	2	.80	-	.91	.95
	3	.68	.75	-	.96
	4	.77	.82	.83	-

Note: Correlation Coefficients are above the diagonals.

[a]All correlation coefficients are significant ($p < .001$)